Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2007**	2. REPORT TYPE **final**	3. DATES COVERED **00-00-2007 to 00-00-2007**

4. TITLE AND SUBTITLE **The knowledge matrix approach to intelligence fusion**	5a. CONTRACT NUMBER **W74V8H-06-C-0001**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **Christopher Pernin; Louis Moore; Katherine Comanor**	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **RAND Corporation,1776 Main Street,Santa Monica,CA,90401-3208**	8. PERFORMING ORGANIZATION REPORT NUMBER **RAND/TR-416-A**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **Captain James Malcom, HQ USAF/A8XP, Room 4D1083, 1070 Air Force Pentagon, Washington, DC, 20330-1070**	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Online access|uhttp://www.rand.org/pubs/technical_reports/TR416/

14. ABSTRACT
As the U.S. military transforms to an information-based force, it will need processes and methods to collect, combine, and utilize the intelligence that is generated by its assets. The process known as fusion will play an important role in determining whether this intelligence is used in the most beneficial manner. The process of fusion, combining pieces of information to produce higher-quality information, knowledge, and understanding, is often poorly represented in constructive models and simulations that are used to analyze intelligence issues. This report describes one approach to capturing the fusion process in a constructive simulation, providing detailed examples to aid in further development and instantiation. The sequential fusion method in intended to determine whether separate intelligence observations are close enough geographically, have consistently identified the same battlefield entity, and contain high-quality information, all of which must be considered before fusion of intelligence can occur. The fusion process described in this report is, for the most part, an implicit representation of the generation of battlefield intelligence and can be used in a constructive simulation or fusion model to better understand the dynamics of intelligence-gathering systems and their effect on intelligence metrics.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**55**	

This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

TECHNICAL
REPORT

The Knowledge Matrix Approach to Intelligence Fusion

Christopher G. Pernin, Louis R. Moore,
Katherine Comanor

Prepared for the United States Army

RAND ARROYO CENTER

The research described in this report was sponsored by the United States Army under Contract No. W74V8H-06-C-0001.

Published 2007 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

This report is part of the second phase of a project titled “Measuring the Value of High-Level Fusion, Intelligence Preparation of the Battlefield, and Robust Operational Command and Control Capabilities for the Army Objective Force.” The project aims to understand the quantitative effects of better command, control, communication, computer, intelligence, surveillance, and reconnaissance (C4ISR) capabilities on battle outcome. This report describes the representation of the fusion process and, specifically, its implementation in a combat simulation. It should be of interest to those concerned with the analysis of C4ISR issues and their representations in combat simulations.

This research was supported through the Army Model and Simulation Office (AMSO) C4ISR Focus Area Collaborative Team (FACT). It was sponsored by the Office of the Deputy Chief of Staff for Intelligence (U.S. Army G-2) and funded by the Army Model Improvement Program (AMIP) and the Army’s Simulation Technology (SIMTECH) program. The research was conducted in RAND Arroyo Center’s Force Development and Technology Program. RAND Arroyo Center, part of the RAND Corporation, is a federally funded research and development center sponsored by the United States Army.

For comments or further information, please contact the authors: Christopher G. Pernin (Chris_Pernin@rand.org) and Louis R. Moore (Louis_Moore@rand.org).

The Project Unique Identification Code (PUIC) for the project that produced this document is DAMIIX001.

For more information on RAND Arroyo Center, contact the Director of Operations (telephone 310-393-0411, extension 6419; FAX 310-451-6952; email Marcy_Agmon@rand.org), or visit Arroyo's Web site at http://www.rand.org/ard.

Contents

Figures

Tables

Summary

As the military transforms into an information-based force, it will need processes and methods to collect, combine, and utilize the intelligence that is generated by its assets. The process known as *fusion* will play an important role in determining whether this intelligence is used in the most beneficial manner to support the military's vision of an information-centric force.

The process of fusion, combining pieces of information to produce higher-quality information, knowledge, and understanding, is often poorly represented in constructive models and simulations that are used to analyze intelligence issues. However, there have been efforts to rectify this situation by incorporating aspects of information fusion into combat simulations. This report describes one approach to capturing the fusion process in a constructive simulation, providing detailed examples to aid in further development and instantiation.

The analytical method proposed here is a sequential process of determining the quality of a piece of information and the likelihood that two pieces of information concern the same entity[1] or, indeed, two separate entities. The process entails putting the observations through a sequence of operations to determine whether they (1) are close enough geographically[2] with respect to their separate errors in location to be of the same entity, (2) have consistent identities that would not prevent them from being considered the same entity, and (3) contain information content of high enough quality to warrant the combination.

Once two observations have passed these three tests, a combination process determines the fused product. In cases in which additional information about an entity is generated—for example, knowledge of the location of a superior unit gleaned from knowing where the subordinate is—the process is able to capture the information in the common operational picture (COP). Higher-level fusion, such as the generation of aggregates, is also captured in the fusion process.

The fusion process provides a means of adding information to the COP and, in doing so, quantifies the quality of individual and combined intelligence, as well as higher-order fusion products. The fusion process described in this report is, for the most part, an implicit representation of the generation of battlefield intelligence and can be used in a constructive simulation or fusion model to better understand the dynamics of intelligence-gathering systems and their effect on intelligence metrics. Where explicit representations of fusion also exist, such as in the case of the generation of location information, both representations are included. The process includes descriptions of stochastic as well as deterministic representations. The representations

1 *Entity* is used loosely in this report. We include in this definition all common entities, such as vehicles and buildings, as well as more nebulous entities, such as events and relationships.

2 *Close enough* can also include temporality for moving objects.

in this report are largely reflective of intelligence fusion in the physical domain; other aspects of the human and information domains (e.g., intent) are included, though in-depth exploration of these is outside the scope of this work.

The approach described in this report is largely reflective of the work of Keithley (2000), and our research group has incorporated it into a stochastic, agent-based simulation to help with the analysis of C4ISR systems and concepts for the Army. However, we describe here much of the method and calculus involved to aid in further development and inclusion into future military simulations.

Abbreviations

AMIP	Army Model Improvement Program
AMSO	Army Model and Simulation Office
ASAS	All Source Analysis System
C4ISR	command, control, communication, computer, intelligence, surveillance, and reconnaissance
CEP	circular error probability
COP	common operational picture
DoD	U.S. Department of Defense
FACT	Focus Area Collaborative Team
GMTI	ground moving target indicator
IPB	intelligence preparation of the battlefield
JDL	Joint Directors of Laboratories
MCO	major combat operation
MTI	moving-target indication
OOB	order of battle
PIR	priority intelligence requirement
SAR	synthetic aperture radar
SIGINT	signals intelligence
SIMTECH	Simulation Technology (program)
SME	subject-matter expert

CHAPTER ONE

Introduction

As the military transforms into an information-based force, it will need processes and methods to collect, combine, and utilize the intelligence that is generated by its assets. The process known as *fusion* will play an important role in determining whether this intelligence is used in the most beneficial manner to support the military's vision of an information-centric force.

The process of fusion, combining pieces of information to produce higher-quality information, knowledge, and understanding, is often poorly represented in the constructive models and simulations used to support the analysis of intelligence issues. However, there have been efforts to rectify this situation by incorporating aspects of information fusion into combat simulations. This report describes one approach to representing the fusion process in a constructive simulation, and it provides detailed examples to aid in further development and implementation. The approach is largely reflective of the work of Keithley (2000), which our research group has incorporated into a stochastic, agent-based simulation to facilitate the analysis of command, control, communication, computer, intelligence, surveillance, and reconnaissance (C4ISR) systems and concepts for the Army. Here, we describe much of the method and calculus involved to aid in further development and inclusion into future military simulations.

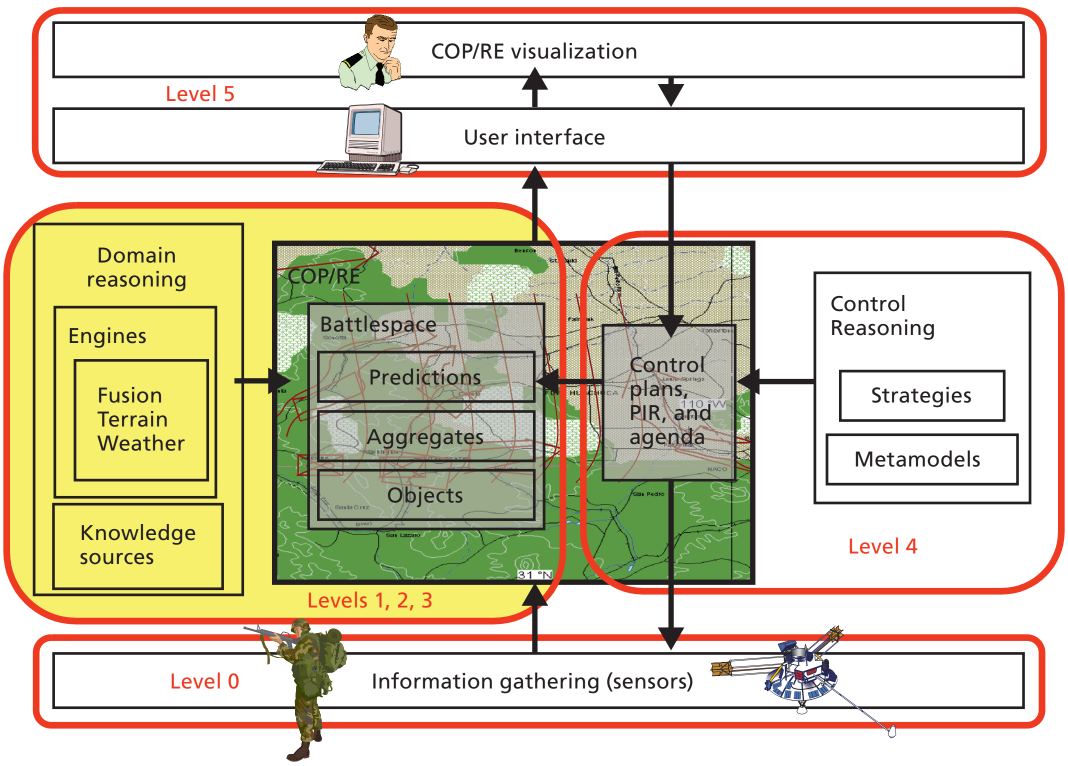
This report describes an implicit representation of the fusion process. We are not proposing a solution to the problem of combining information from multiple sensors or across multiple time steps that may represent incomplete, incoherent subsets of data. Thus, our representation does not attempt to replicate the actual method used in military intelligence systems, nor does it aim to explicitly model every facet of the actual fusion process. We are proposing a characterization of a process that may be incorporated into a simulation. This report is intended to provide a platform to further the thinking of the fusion process as a whole.

Our representation, therefore, incorporates many of the specific aspects of intelligence generation, fusion, and outputs and measures of effectiveness of the process, but without having to explicitly represent the actual generation and fusion of intelligence. The advantage of this representation of fusion relative to explicit representations of fusion lies in its ability to represent many diverse fusion processes and in its execution speed. It is driven by and heavily dependent on data describing the quality of the output of the underlying fusion process. Thus, our approach is only as good as the underlying data. The representations in this report are largely reflective of intelligence fusion in the physical domain; other aspects of the human and information domains (e.g., intent) are included, though in-depth exploration of these is outside the scope of this work.

Figure 1.1 shows the levels and the interrelationships among the various fusion levels.[1] The Joint Directors of Laboratories (JDL) Data Fusion Model was first proposed in 1985 under the guidance of the U.S. Department of Defense (DoD).[2] The model partitioned fusion into six (not necessarily sequential) levels that described the different products and activities associated with the use and manipulation of intelligence data. The JDL model is a functional model that clarifies many of the fusion processes currently being discussed and implemented in the military intelligence community. It has evolved and expanded over the years to accommodate the improved understanding of the data-fusion processes necessary to manage joint intelligence requirements.

In the JDL fusion model, level 0 involves the organization and normalization of the data in an observation. Levels 1, 2, and 3 correlate, aggregate, and interpret observations, and they are the main emphasis of this report. Level 4 assesses and controls the fusion process to drive improvements in the process. It also allocates resources to satisfy collection needs. Level 5 provides the process with customer feedback and control through visualization of the fusion

Figure 1.1
Joint Directors of Laboratories Data Fusion Model



NOTE: COP = common operational picture. PIR = priority intelligence requirement. RE = running estimate.
RAND *TR416-1.1*

---

1 Various models have been created to aid and describe data fusion. See, for example, Dasarathy (1997), Bedworth and O'Brien (2000), Salerno (2002), Endsley (1995, 2000), Blasch and Plano (2003), and U.S. Army Intelligence Center (2004).

2 The original JDL model included only levels 1 through 4; levels 0 and 5 were subsequently added.

products and determination of PIRs. The last two levels are beyond the scope of this project, though they remain a challenge to practitioners.

The method described in this report is concerned with fusion levels 0 through 3. The first, level 0, organizes data into forms that can be used by the process. In our case, we will focus on sensor data that are produced on the battlefield. However, the process and implementation are not contingent on this. The data are assumed to come from a variety of sources, whether a sensor on a platform; a person seeing, reading, or hearing something; or some other source in the battlespace, all of which can be incorporated into the operational picture. In the next chapter, we describe the generation of the data from an observation and how the data are represented implicitly through the use of a knowledge matrix. Later chapters describe the subsequent three levels of fusion.

CHAPTER TWO

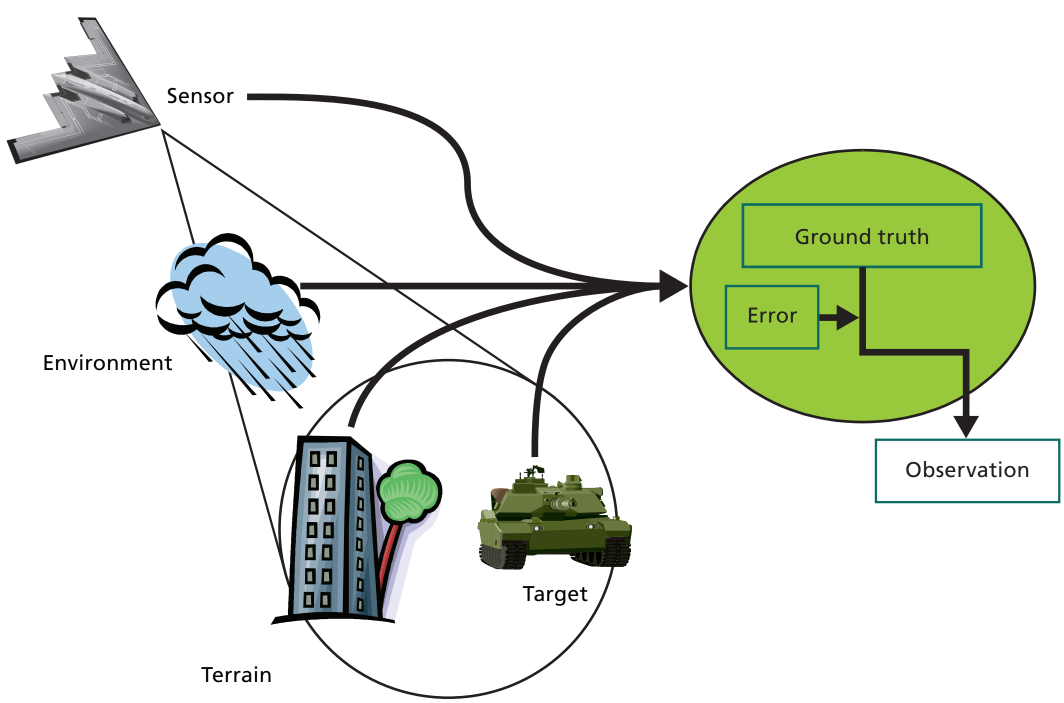
The "Knowledge Matrix"

This report walks through three areas of the fusion process: the generation of an observation, the combination of two (or more) observations, and the representation of higher-level fusion. These steps in the fusion process are described with examples and some of the specific mathematical formulations embedded in the representations. This report also describes both stochastic and deterministic routes to implementing the method.

First, however, we address the issue of how to describe the quality of an observation. Factors affecting observation quality are shown in Figure 2.1.

Several factors contribute to the quality of an observation. The environment and terrain directly impact whether a sensor is able to observe a given object on the battlefield and how good that observation is. The amalgamation of the factors involved in observing an object can be described loosely as the error in the perception. The quality of the observation is

Figure 2.1
Some Factors Affecting Observation Quality



RAND *TR416-2.1*

a function of how close the perception is to ground truth: In a high-quality observation, the perception of the object is close to ground truth, and in a low-quality observation, the object perceived is far from ground truth.

Four types of battlefield entities are considered and represented in this fusion model:

- infrastructure and facilities (e.g., buildings, roads, bridges)
- pieces of equipment (e.g., tanks, trucks)
- aggregates (e.g., units, collections, organizations)
- structured relationships (such as an order of battle [OOB]).

In addition, two types of entities are modeled for non–major combat operation (MCO) intelligence:

- significant events occurring
- persons forming a network.

The first three—infrastructure and facilities, pieces of equipment, and aggregates (collections of the previous two types)—are easily understood. The fourth type (structured relationships) and the two others now being developed (namely, events and networks) are much more abstract. They have to do with higher levels of information that cannot be sensed in a traditional sense but may require analysis to identify. Representing the fourth type of entity in intelligence-collection methods and fusion schemes affords an opportunity to go beyond analysis of operations other than MCOs to include the full spectrum of operations (Headquarters, U.S. Department of the Army, 2001). Relationships, events, and networks are not directly discussed in this report.

One method of quantitatively characterizing the quality of an observation is the approach described by Keithley (2000). The knowledge matrix approach quantifies the quality of an observation according to six knowledge types: location (Loc), identification (ID), track, activity (Act), capability (Cap), and intent. The levels range from low to high quality, and each level has associated with it a probability that the observation meets or exceeds a given level. This information is collected into a table called the *knowledge matrix*.

Table 2.1 illustrates this quantitative approach to observation quality.

Table 2.1
Quantitative Assessment of Observation Quality as a Probability of Achieving Quality Levels

Quality Level	Loc	ID
5 (high)	0.10	0.70
4	0.20	0.80
3	0.60	0.90
2	0.80	0.95
1	0.90	0.99
0 (low)	1.00	1.00

In Table 2.1, only two columns are shown: those for location and identification. More columns can be added as necessary. In each of these two columns are the likelihoods that an observation will meet or exceed the quality listed in the leftmost column. The qualities range from low to high. In this particular example, the likelihood of meeting or exceeding a quality level of high for each location and identification type is 0.10 and 0.70, respectively, meaning that it is likely that this observation produced high-quality knowledge of identification and rather low-quality knowledge for location.[1]

Table 2.2 gives examples of qualitative (identification) and quantitative (location) metrics for describing the varying levels of knowledge quality. Each of the qualities can have either a qualitative or a quantitative description. For example, a low-quality observation of an object might indicate only that detection occurred, with little information about the object that was detected. As the quality improves (moving from bottom to top in Table 2.2), more information about the object or entity is obtained. For example, the information distinguishes whether the object is a vehicle or a structure (quality level 2); if it is a vehicle, whether it is wheeled or tracked (quality level 3); and finally, at the highest quality level, what the object is, as well as its parent or group membership. We note that, if an observation of a given quality has been obtained, it naturally implies that information associated with the lower quality levels has been obtained as well. In other words, if an object can be classified as wheeled or tracked (quality level 3), it is also known whether this object is a vehicle or a structure (quality level 2).

A quantitative description of quality level is also possible. In the case of knowledge regarding location, the quality level might correspond to location errors. A low-quality observation might be a location error of 10 km (quality level 0), whereas a high-quality observation might be a location error of 5 m (quality level 5).

Keithley (2000) developed descriptions, shown in Table 2.3, of the quality levels associated with six different knowledge types: location, track, identification, activity, capability, and intent. The qualities for location are continuous and quantitative. In comparison, for the remaining types of knowledge, the quality descriptions are qualitative and discrete.

Sensor capabilities can be captured in the knowledge matrix. The knowledge types and associated likelihoods of reaching a given quality level can be indicative of the type of sensor

Table 2.2
Metrics for Describing Levels of Knowledge Quality

Quality Level	ID	Loc
5 (high)	Specify object and parent	5 m
4	Specify object	10 m
3	Classify (e.g., wheeled or tracked vehicle)	20 m
2	Distinguish (e.g., vehicle or structure)	100 m
1	Discriminate	1 km
0 (low)	Detect	10 km

[1] The reader should keep in mind that these levels are not necessarily precise and are left to the analyst to determine what is appropriate.

Table 2.3
Qualitative and Quantitative Descriptions of the Quality Levels for Each Knowledge Type

	Knowledge Type					
Quality	**Loc**	**Track**	**ID**	**Act**	**Cap**	**Intent**
Highest 5	5 m	Vectors and patterns	Specify object and parent	Precise actions	All elements	All long- and short-term objectives
High 4	10 m	Vectors	Specify object	Many specific actions	Many details	Major objectives
Medium 3	20 m	General speed and direction	Classify (e.g., wheeled, tracked vehicle)	Identifiable actions	Some details	Primary objectives
Medium-low 2	100 m	Toward or away	Distinguish (e.g., vehicle, structure)	Single identifiable action	General information	General objectives
Low 1	1 km	Stationary or not	Discriminate	Unidentifiable action	Minimal information	Single objectives
Lowest 0	10 km	Detect	Detect	Detect	Detect	Detect

and target being observed. For example, as shown in Table 2.4, a synthetic aperture radar (SAR) sensor might have a particularly good identification capability (i.e., a 0.70 probability of reaching high-quality identification knowledge in the example on the left) and a rather low probability of providing a track of an object (in this case, a 0.0 probability of anything other than the lowest quality for track). This capability can be compared with that of a ground moving target indicator (GMTI) sensor, which might be just the opposite: very high probability of tracking an object and low probability of identification. The exact likelihoods in the knowledge matrixes may be determined primarily by the type of sensor and target being observed, but they will also depend on the terrain and environmental factors present when a given observation occurred.

Table 2.4
Capturing Sensor Capabilities in a Knowledge Matrix

	SAR Sensor		GMTI Sensor	
Quality	**Track**	**ID**	**Track**	**ID**
Highest	0	0.70	0	0.00
High	0	0.80	1	0.30
Medium	0	0.90	1	0.80
Medium-low	0	0.95	1	0.85
Low	0	0.99	1	0.90
Lowest	1	1.00	1	1.00

Generating an Observation: Level 0 Fusion

Having established how to describe an observation across multiple types of knowledge, we now turn to the process of generating an observation as is done in a combat model or simulation, illustrated in Figure 2.2. The generation of the observation is part of level 0 and level 1 fusion, as described earlier.

Three key factors affect the quality of a sensor's observation of a given target: environmental effects, such as weather and clutter; terrain effects; and the type of sensor used to generate the observation.[2] A sensor can generate various types of knowledge about a target, such as its location and its identity. The quality of this knowledge will vary from one instance to the next, depending on the factors involved. In each particular case, the quality of a sensor's observation is captured in its knowledge matrix.

The quality of a particular sighting can be deduced from its corresponding knowledge matrix. The deduction can occur in one of two ways. In a stochastic simulation, the observation quality for a particular knowledge type is generated from a random variable (U) uniformly distributed between 0 and 1. The quality reported for that instance is the highest quality with a likelihood greater than U.

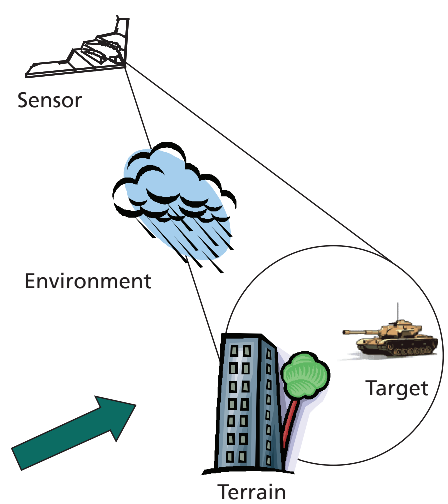
Table 2.5 gives an example for a single knowledge type, identification (*ID*). If the uniformly distributed random variable[3] for identification (U_{ID}) is 0.23, which is greater than 0.0 and not greater than 0.3, the observation would generate an identification quality level of 4. This quality level corresponds to knowing the "object specification." Since attaining a

Figure 2.2
Overview of the Generation of an Observation

Each observation has an associated knowledge matrix…

Quality	Loc	ID	...
5	*0.0*	*0.0*	...
4	*1.0*	*0.30*	...
3	*1.0*	*0.80*	...
2	*1.0*	*0.85*	...
1	*1.0*	*0.90*	...
0	*1.0*	*1.0*	...

…that is based on sensor coverage patterns, probability of detection, weather, foliage, target activity, and other factors.



RAND *TR416-2.2*

[2] Others include enemy countermeasures, such as jamming and deception.

[3] To generate a random variable in the model, we started with a uniform distribution and inverted the distribution function (which is contained in the knowledge matrix).

Table 2.5
Example Portion of a Knowledge Matrix

Quality	ID
5	0.00
4	0.30
3	0.80
2	0.85
1	0.90
0	1.00

quality level of 4 indicates that lower quality levels are also known, the observation has indicated whether the object is a vehicle or a structure, and (if it is a vehicle) whether the object is tracked or wheeled, in addition to what specific object it is. With a quality level of 4, the observation might have indicated that the object is a T-72 (e.g., a vehicle, tracked, Russian T-72).

If $U_{ID} = 0.84$, which is greater than 0.8 and not greater than 0.85, the observation would have generated a quality level of 2 for identification. In this case, only whether the object was, for example, a vehicle or a structure could have been deduced from the sighting.

The second way to deduce the actual quality level of information in a given knowledge matrix is to calculate the mean of each column.[4] The mean quality of a knowledge category is the sum of the column entries in the category minus 1. This number is rounded up to determine the quality level reported. In the example shown in Table 2.5, the average of the identification column is 2.85, which, rounded up, equates to a quality level of 3. In this case, whether the object, if it is a vehicle, is wheeled or tracked could have been deduced from the observation. This particular method of determining the reported quality from a given knowledge matrix is useful in a deterministic simulation in which random variables are not used.

The knowledge matrix method can also capture implicit representations of the quality of the knowledge types that may be associated with a given observation. In some cases, it may be possible to generate explicit representations of knowledge. One example of this is the representation of location.

Stochastic simulations typically apply errors to ground truth locations of the form

$$(x, y) = (x_{gt}, y_{gt}) + (Z_x\sigma_x, Z_y\sigma_y),$$

where (x_{gt}, y_{gt}) is the ground truth location, Z_x and Z_y are standard normal variables, and σ_x and σ_y are the standard deviations of the location error.

The reported location (perception) is (x, y), and the perceived standard deviations are $(\sigma_x$ and $\sigma_y)$, which may not equal the actual standard deviations. A deterministic simulation might report the actual ground truths, $(x_{gt}$ and $y_{gt})$, along with the standard deviations,

[4] Use of the median can be considered, but no analysis has been done to determine which is better (or whether either is, in fact, more appropriate).

(σ_x and σ_y). Various specific processes for location fusion have been documented in the literature on the subject.[5]

Once the new locations and standard deviations are explicitly calculated in the simulation, it is possible to generate the knowledge matrix from the explicitly generated observation. The probability that the distance between the observed location and ground truth location is less than d is given by

$$1-\exp\left(\frac{-d^2}{2\sigma^2}\right),$$

where

$d=[(x-\sigma)^2+(y-\sigma)^2]^{0.5}$

$\sigma=$ ground truth location, assuming that

$\sigma_x=\sigma_y=\sigma.$

Two examples are shown in Table 2.6. In example 1, the knowledge matrix has been calculated for an explicit standard deviation of 10; in example 2, the standard deviation is 100. As expected, the knowledge matrix in example 1 has a much higher probability (0.12) of achieving a quality level of 5 than the knowledge matrix in example 2 (0.001). The likelihood of achieving other quality levels is similarly skewed.

Table 2.6
Two Examples of Knowledge Matrix Computation for Location from Explicitly Generated Observations

		Example 1: $\sigma=10$	Example 2: $\sigma=100$
Quality	***d***	**Loc**	**Loc**
5	5	0.12	0.001
4	10	0.39	0.005
3	20	0.87	0.020
2	100	1.00	0.390
1	1,000	1.00	1.000
0	10,000	1.00	1.000

[5] See Washburn (2004) for a discussion of a more general process.

Combining Two Observations: Level 1 Fusion

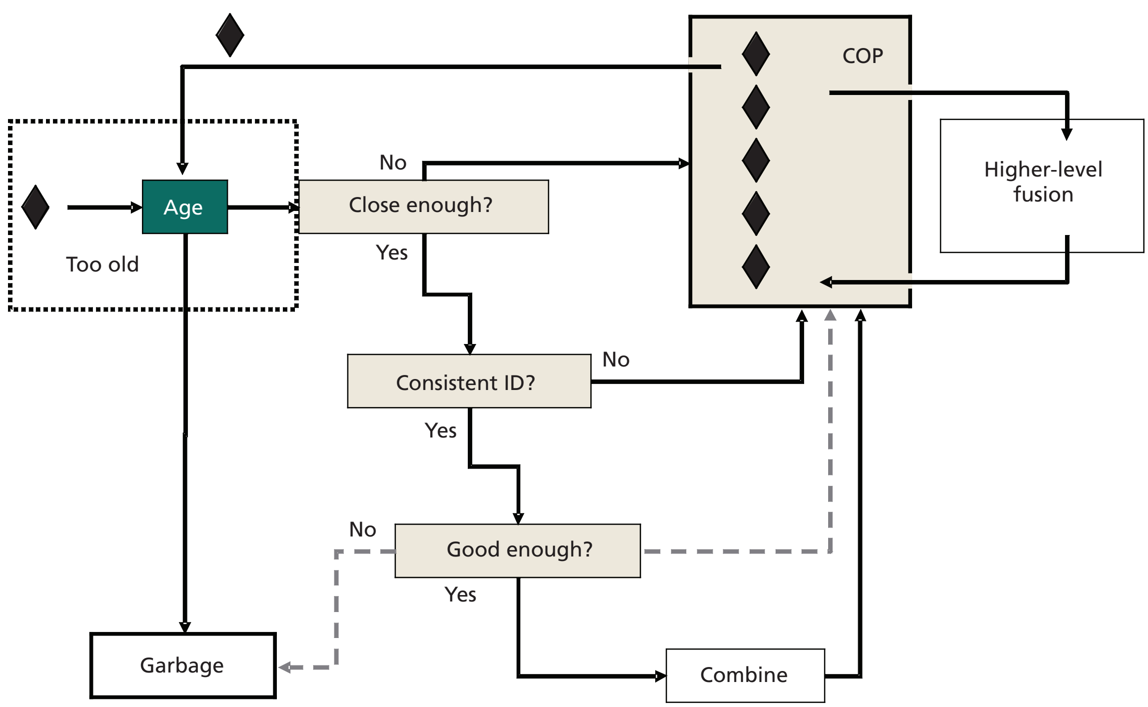
Now that an observation has been generated and an associated knowledge matrix and quality level have been established, we turn to combining multiple observations. The combination of multiple sensor outputs conforms to level 1 fusion, wherein we have identified that discrete entities or events have been observed, and it is time to correlate and combine information and resolve any conflicts in the data.

A series of steps determines whether two observations should be combined. The first step involves aging the new observations (as well as those that have already been processed on the COP) and expunging observations that are deemed too old. This process is illustrated in Figure 2.3. It should be noted that each intelligence domain executes such a series of steps in parallel with other domains, producing a COP for each domain. This is known as *single-source processing*. Then, an integrating sequence, *all-source processing*, unifies the separate COPs into one.

Aging the COP

Information becomes less accurate over time. If an object moves or if an event changes the circumstances, something that may have been true at one time becomes less certain later. To that end, it is important to incorporate a method for aging the knowledge matrix based on a set of circumstances. Some types of information may age more quickly than other types. The accuracy of a bit of information will depend not only on the target but also on environmental effects. In this section, we explain one method of incorporating this effect into the knowledge matrix approach.

Figure 2.3
Steps Involved in Populating the COP: Age



NOTE: Diamonds signify observations.
RAND *TR416-2.3*

The aging method we use relies on the concept of data half-life—namely, the time it takes for half the quality of the information to be lost (see Table 2.7 for examples). To calculate the value of the aged knowledge matrix entry, G_{Aged}, we must determine the decay constant, R. The half-life time for each type of OOB object is calculated through expert opinion and corresponds to the time it takes for the knowledge of the sighting to decay to one-half its original value. Representative values for three OOB objects—foot mobile, vehicular, and fixed-wing aircraft—are shown in Table 2.7. We define exponential decay as

$$G_{Aged} = G_0 e^{-R \times \Delta t}.$$

From this equation, we see that $R \times \Delta t = 0.693 = \ln(2)$ when Δt equals the half-life of the data. We know Δt for each object and therefore can calculate *R*. The results are shown in Table 2.7.

Table 2.8 shows an example of the aging process for a 10-minute-old ground target. The original matrix is on the left; the aged-observation matrix is on the right. Values are shown for location and identification. We see that a quality level of 3 for location and identification has been reduced over a period of 10 minutes to quality levels of 2 and 0, respectively. After 10 minutes, the information that was available on its identification is rendered almost completely

Table 2.7
Half-Lives and R-Values for Some Vehicle Types

Vehicle Type	Speed (km/hr)	Half-Life (min)	*R*
Foot	1–4	30	0.023
Ground	5–49	10	0.069
Aircraft	1,000	0.5	1.386

Table 2.8
Location and Identification Knowledge Matrix Entries for a Ground Target

	Ground-Target Observation		Aged Ground-Target Observation	
Quality	**Loc**	**ID**	**Loc**	**ID**
5	0.06	0.42	0.03	0.21
4	0.13	0.73	0.06	0.37
3	0.25	0.90	0.12	0.45
2	0.90	0.96	0.45	0.48
1	0.99	0.98	0.49	0.49
0	0.99	0.99	0.49	0.49
Uniform	0.20	0.84	0.20	0.84

NOTE: Values are given both at inception ("Ground-Target Observation") and after being aged for 10 minutes ("Aged Ground-Target Observation"). The reported quality level generated from a uniform random draw ("Uniform") is indicated by shading. The uniform random draw stays the same throughout the aging process.

useless. Note that, in each of these cases, the uniform random variables listed below the matrix are unchanged after aging.

Observations should be dropped from the COP when they are too old. An observation that has sat on the COP and not been updated or changed for a certain period of time may be deemed unnecessary. In this case, each observation or class of observations may be given a maximum time on the COP before it has to be either updated or dropped.

Aging the knowledge matrix is an implicit means of determining the effects of time on an observation's quality. Some of the knowledge columns in the matrix, however, may be amenable to explicit calculation of the effects of time on quality. For example, the location can be aged by updating the position based on the initial location and velocity, assuming a straight path (better known as *dead reckoning*).[6] In this case, given the position, velocity, and associated errors in each, we can explicitly calculate an expected position at some time in the future. To determine the dead-reckoning location after Δt time, the location (x, y) is

$$(x_{gt}, y_{gt}) + (Z_x\sigma_x, Z_y\sigma_y) + \Delta t\left[(\Delta x_{gt}, \Delta y_{gt}) + (Z_{\Delta x}\sigma_{\Delta x}, Z_{\Delta y}\sigma_{\Delta y})\right],$$

where (x_{gt}, y_{gt}) is the ground-truth location; $(\Delta x_{gt}, \Delta y_{gt})$ is the ground-truth velocity; Z_x, $Z_y, Z_{\Delta x}$, and $Z_{\Delta y}$ are standard normal variables; σ_x and σ_y are the standard deviations of the location error for (x, y); and $\sigma_{\Delta x}$ and $\sigma_{\Delta y}$ are the standard deviations of the velocity error for (x, y).

In the case of knowledge about a potential track, an offline model might be used to determine average (or the distributions of) fused track durations and errors that can be applied directly to knowledge quality (Steinberg and Rosen, 1997). We do not discuss such a model here.

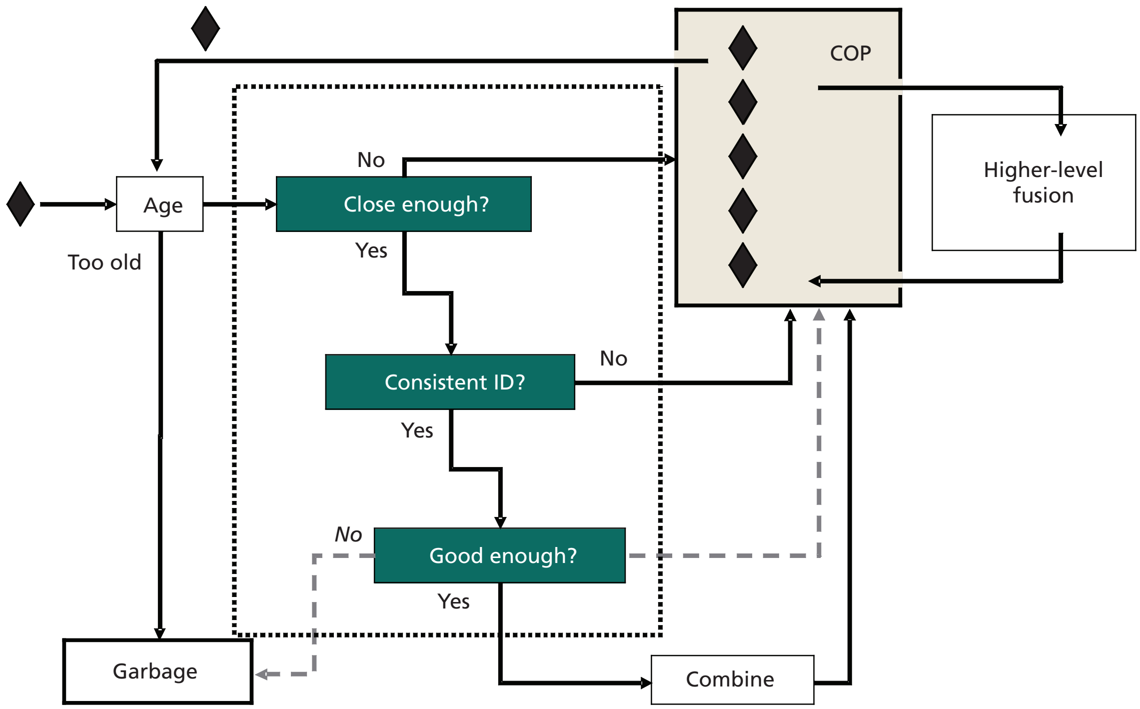
Once an observation has been aged, we determine whether it will be combined with an observation already on the COP.[7] The information on the COP should similarly be aged, either explicitly or implicitly, before any combination criteria are assessed. (Intelligence preparation of the battlefield [IPB] information, which is discussed later, is handled during this process in a slightly different manner from that for handling standard observations on the COP.) The process starts by taking a new observation and comparing it with an observation on the COP. There are three criteria that need to be assessed to determine whether the two observations should be combined. They are to check whether the observations are close enough, whether they have consistent identifications, and whether the knowledge matrixes are good enough, as illustrated in Figure 2.4.

Figure 2.5 illustrates the necessary fusion conditions that must be met before two observations can be combined. It shows a sample knowledge matrix for each of two observations. Beneath each knowledge matrix is a set of numbers (uniformly distributed random variables, U_i), which are used to generate the perception. As can be seen, we have two observations with two reported locations and standard deviations based on an explicit representation of location knowledge. (Because the location was explicitly generated, the uniform random variable

[6] We are using the limited, though oft-used, *dead reckoning* as an example of an explicit movement capture. Other methods might be integrated and just as easily incorporated into the overall knowledge matrix representation.

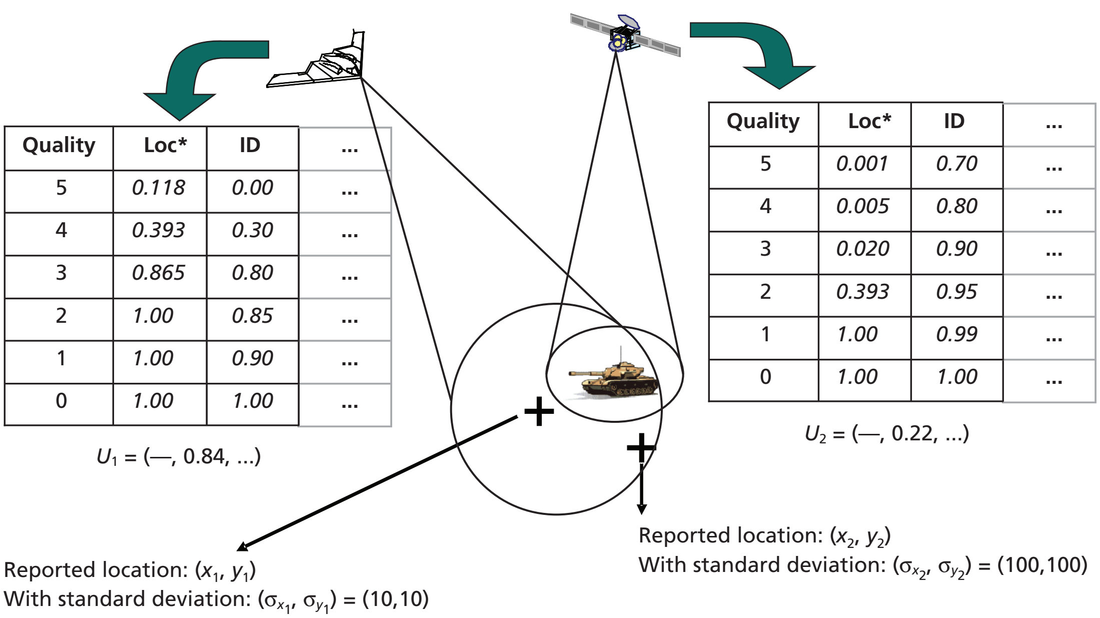
[7] Combining the observations is typically referred to as the process of *correlation* in standard fusion terminology.

Figure 2.4
Steps Involved in Populating the COP: Compare Observations



NOTE: Diamonds signify observations.
RAND *TR416-2.4*

Figure 2.5
Generation of Two Observations of One Target



Quality	Loc*	ID	...
5	*0.118*	*0.00*	...
4	*0.393*	*0.30*	...
3	*0.865*	*0.80*	...
2	*1.00*	*0.85*	...
1	*1.00*	*0.90*	...
0	*1.00*	*1.00*	...

Quality	Loc*	ID	...
5	*0.001*	*0.70*	...
4	*0.005*	*0.80*	...
3	*0.020*	*0.90*	...
2	*0.393*	*0.95*	...
1	*1.00*	*0.99*	...
0	*1.00*	*1.00*	...

*Assumes explicit representation of location knowledge.
RAND *TR416-2.5*

is not generated and is replaced in the picture with a hyphen.) The case on the left has one-tenth (standard deviation = 10) the standard deviation of the sighting on the right (standard deviation = 100). Only two columns of the knowledge matrix are shown, for the sake of simplicity.

There is no reason to assume that the two example observations are of a single entity on the ground (in these examples, a tank). The two observations could easily be of two or more separate entities on the battlefield. The criteria for combining the observations will include heuristics to determine whether the observations are of the same entity. In addition, it may be the case that a single sensor observed one or multiple entities on the ground to generate the two observations being considered (an example later in this report addresses this scenario). In that case, once again, the following criteria can be used to determine whether the observations are of the same entity and should be combined or whether they are from different entities and should remain separate in the operational picture. That a certain sensor or set of sensors made the initial observation will drive only the order in which observations are considered for combination. In general, single-source fusion is performed before multisource fusion. However, the specific logic behind the choice in order, while important, is typically left to subject-matter experts (SMEs).

Are the Observations Close Enough?

The first test to determine whether two observations should be combined is to check whether they are close enough. This entails using information about the error in the locations and relating it to how far apart the two sightings are. The observations should be close enough relative to the errors in the sighting to be considered for combination. The standard deviations in location are assumed, in this case, to be explicitly generated in the simulation. To determine whether the two observations are close enough, we check whether the square of the standardized distance is less than the critical value of a chi-squared random variable with two degrees of freedom (Zelen and Severo, 1972, pp. 940–943; Kenney and Keeping, 1951, pp. 98–100).[8]

Are the Observations Consistent?

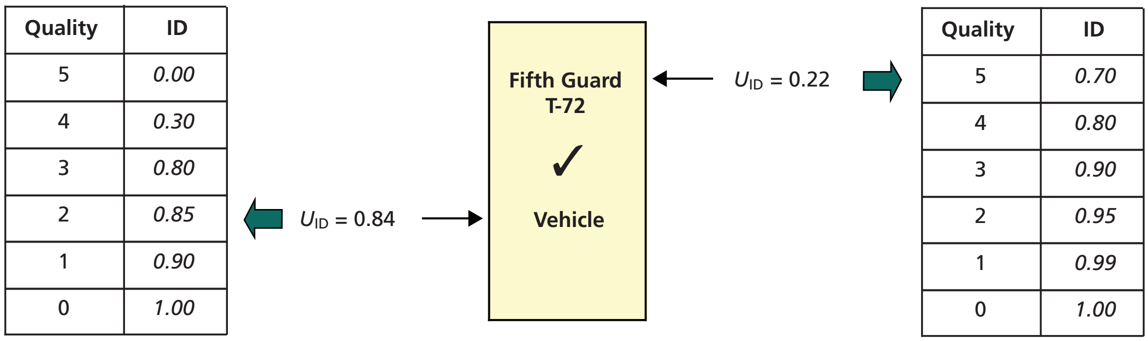
The second criterion that is evaluated when considering the combination of two observations is whether the information on the identity of the entities is consistent. Consistency does not require that the perception of identity be the same for both observations but, rather, that the perception of identity does not conflict. For example, perceiving one entity to be a vehicle, but not knowing what type of vehicle, would be consistent with an observation that indicated that the entity was a T-72 (a specific type of vehicle).

In the example in Figure 2.6, the observation on the left (with a uniform random variable of 0.84) produces information indicating that the entity is a vehicle. The observation on the right (with the uniform random variable of 0.22) indicates that it is a T-72 tank from the Fifth Guard. While the observation on the left has considerably less information about the identity of the object, it is consistent with the observation on the right because it does not contradict the knowledge that it is a specific tank from a specific unit—a T-72 from the Fifth Guard is also a vehicle.

---

[8] In the examples in this report, we use a 95-percent confidence level for the chi-squared test. This can be set by the user, depending on tolerance for redundant observations.

Figure 2.6
Example of Consistency of Two Identities



NOTE: The two observations are consistent but not identical.
RAND *TR416-2.6*

In the absence of any ambiguity in misidentification or other effects that may cause a false identification (and not just an identification of lower quality), individual observations of a specific entity will always have consistent knowledge of identity and, therefore, will always pass this test.

Are the Observations Good Enough?

The third criterion that must be measured before two observations can combine is the adequacy of the quality of the observation itself. This entails calculating a quality score for each knowledge type and an overall quality score, which indicates the average overall quality of a given observation.

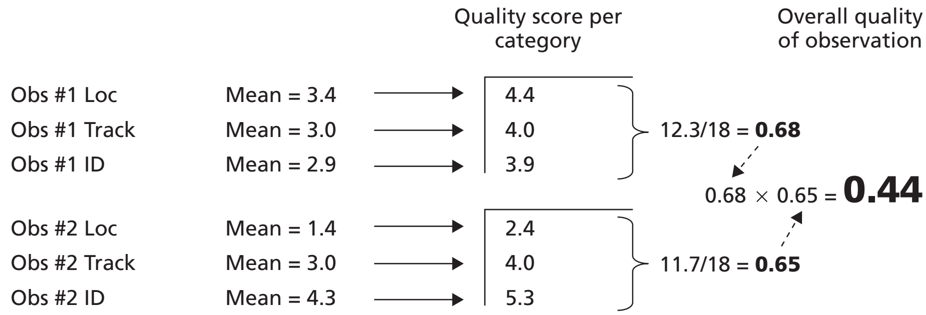
For all knowledge types (i.e., columns of the knowledge matrix), the quality score is $\frac{1}{6}$(mean quality for the type $+$ 1). The overall quality of the observation is the average of all the individual quality scores calculated from the columns of the knowledge matrix. This score is multiplied by the score of another observation, and the resulting fraction is considered the likelihood that the two observations are of the same entity on the battlefield. In a stochastic simulation, this probability can be used as the likelihood that the two observations can be combined. In a deterministic simulation, this number can be used in an absolute manner: If it is above 0.5, the two observations are combined; if it is below, the two observations are considered too poor to combine.

Figure 2.7 shows the individual and overall qualities of two observations for three knowledge types (location, track, and identification).[9] For each observation, the quality score is calculated and a combined average over all knowledge types is determined. In this case, the three mean qualities of the first observation (4.4, 4.0, 3.9) are averaged over the 18 knowledge matrix entries to generate an overall quality of 0.68. This can be multiplied by the result from the second observation (0.65, in this case) to get the probability that the two observations would be good enough to be combined. This method will bias good observations being combined with other good observations and limit the number of bad-overall-quality observations from being combined.

---

[9] These three are typically used for correlating observations.

Figure 2.7
Observation Qualities Must Be Good Enough



RAND *TR416-2.7*

There are alternative approaches to generating quality scores from an individual observation. The location quality can be calculated as the minimum of 1 and $\frac{1}{3}$(mean quality + 1). This alternative scoring technique will favor observations with good target-location quality and poor identification quality. It creates a quality score that takes into consideration different types of sensors, for example. To favor the observations generated by an moving-target indication (MTI) sensor over a signals intelligence (SIGINT) sensor, one would want a scoring technique that favors observations with good target-location quality and poor identification quality so as not to inappropriately degrade the quality score.

When calculating whether two observations should be combined in the manner described here, some columns may not be used in certain circumstances. If one sensor is known not to have a capability in one knowledge type, it can be assumed that it will not match or be in line with another sensor that does. To ensure that the sensors are judged on their own capabilities, certain columns may be left out when generating the overall quality score. This is easily seen by considering an observation from a SAR sensor. The information generated from a SAR image is not expected to produce reasonable-quality intelligence on tracks, and, indeed, an analyst looking at a SAR image will not expect to find information on a moving target. The quality of the other columns—in this case, knowledge of identity—is the focus of the intelligence. When the overall quality score is calculated in this step, the column for track would be left out of the calculation so as not to degrade what is otherwise a good piece of intelligence.

An alternative scoring method is to base the quality score on the percentile rather than the mean of the column distribution. This may be considered a risk evaluation wherein a risk-averse evaluation might take the 90th percentile quality and use that as the score. Alternatively, any percentile, such as the median or 50th percentile, may be used.

Regardless of which scoring technique is used, the output of this criterion is that either the two observations are combined (discussed next) or one of them—usually the one with lower overall quality—is discarded. The motivation behind expunging the worse observation is as follows. The two observations at this point have been deemed to be close enough and have consistent identities. At this point, we are implicitly evaluating whether the observations are any good to the analyst in maintaining a good overall operating picture. If the quality is quite low compared to that of the current observation on the COP, there is no need to combine them and potentially reduce the overall quality of the observation or add unnecessary uncertainty to the current picture. If they are not being combined, the process might as well discard the observation. In the example shown in Figure 2.7, if a uniform random variable did not indicate

that the two observations should be combined, the second one (quality score = 0.65) would be discarded.

Combination of Two Observations

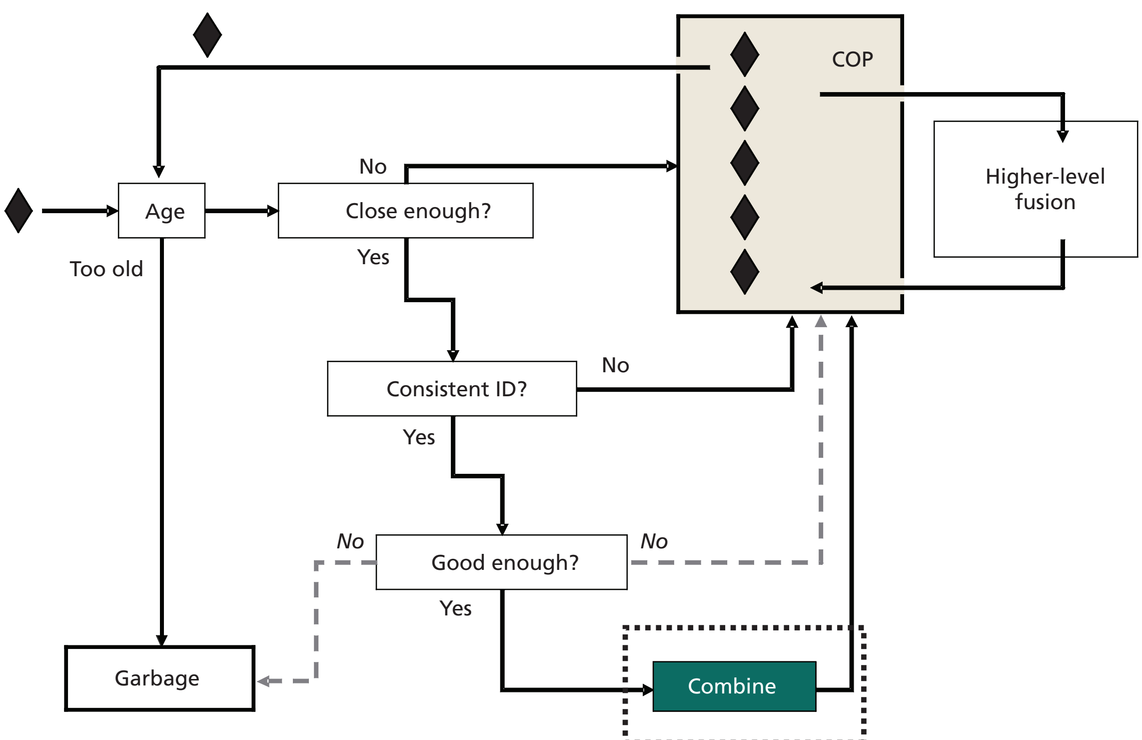
If all three fusion criteria are satisfied, it is possible to combine the two observations into a single observation (see Figure 2.8). This involves two steps. The first is to combine the two knowledge matrixes, and the second is to combine the uniform random variables (U) that are associated with each observation (in the case of the stochastic simulation).

Once the observations have been aged appropriately, and if all three criteria have been met, the two observation matrixes can be combined (see Figure 2.9). The combined matrix, K, formed from observation matrixes G and H, is given by

$$K_{i,j} = 1-(1-G_{i,j})(1-H_{i,j}),$$

where knowledge matrixes G and H are taken from the two observations and K is the new matrix. It has been suggested that this equation is an extension of the Dempster-Shafer theory of evidence (Shafer, 1976). Alternatively, this equation may be derived by assuming independence between the two observations and computing the distribution of the maximum quality of the two.

Figure 2.8
Steps Involved in Populating the COP: Combine Observations



NOTE: Diamonds signify observations.
RAND *TR416-2.8*

Figure 2.9
Combining Two Observations into a Combined Observation Matrix

Likelihood of quality of combined observation is at least as great as that of either original observation.

First observation (G)

Quality	Loc*	Track	ID	Act	Cap	Intent
5	*0.118*	*0.00*	*0.0*	*0.00*	*0.00*	*0.00*
4	*0.393*	*0.00*	*0.30*	*0.00*	*0.00*	*0.90*
3	*0.865*	*0.00*	*0.80*	*0.00*	*0.00*	*0.90*
2	*1.0*	*0.70*	*0.85*	*0.00*	*0.00*	*0.90*
1	*1.0*	*0.90*	*0.90*	*0.00*	*0.90*	*0.90*
0	*1.0*	*0.95*	*1.0*	*1.0*	*0.92*	*0.90*

Second observation (H)

Quality	Loc*	Track	ID	Act	Cap	Intent
5	*0.001*	*0.00*	*0.70*	*0.30*	*0.00*	*0.00*
4	*0.005*	*0.00*	*0.80*	*0.40*	*0.00*	*0.00*
3	*0.020*	*0.00*	*0.90*	*0.50*	*0.70*	*0.00*
2	*0.393*	*0.70*	*0.95*	*0.60*	*0.80*	*0.30*
1	*1.00*	*0.80*	*0.99*	*0.70*	*0.90*	*0.40*
0	*1.00*	*0.90*	*1.00*	*0.80*	*0.92*	*0.50*

Combined observation (K)

Quality	Loc*	Track	ID	Act	Cap	Intent
5	.121	0.00	0.70	0.30	0.00	0.00
4	.403	0.00	0.86	0.40	0.00	0.90
3	.873	0.00	0.98	0.50	0.70	0.90
2	1.00	0.91	0.99	0.60	0.80	0.93
1	1.00	0.98	1.00	0.70	0.99	0.94
0	1.00	1.00	1.00	1.00	0.99	0.95

Where $K_{i,j} = 1 - (1 - G_{i,j})(1 - H_{i,j})$

*Using explicit representation.
RAND *TR416-2.9*

Deterministic simulations are finished at this point, since they will then have a new knowledge matrix and associated mean quality for each knowledge type. In a stochastic simulation, it is necessary to also combine the uniform random variables (U) associated with the knowledge matrix for each observation. The uniform random variables are not redrawn for the new matrix. One approach to combining the random variables is to choose the random variable for each type of knowledge that yielded the maximum quality level for that type. Another is to combine the two random variables as follows. If $U(G)_i$ and $U(H)_i$ are the uniform sampling vectors associated with the knowledge matrixes G and H, respectively, the fused vector $U(K)_i$ is defined as

$$U(K)_i = 1 - \left[1 - U(G)_i\right]\left[1 - U(H)_i\right]\left[1 - \ln\left([1 - U(G)_i][1 - U(H)_i]\right)\right]$$

The advantage of the second approach is that the distribution of the resulting random variables is still uniform.[10] Continuing the previous example, Table 2.9 presents the combined $U(K)$.

Some types of knowledge may be explicitly calculated. For locations in particular, through such means as Bayesian updating, it is possible to calculate errors and new locations from two observations.

[10] This combination is drawn from the fact that, if a continuous random variable, x, has distribution function, F, then $F(x)$ is a uniform random variable.

Table 2.9
Example Combination of the Uniform Random Vectors

Observation	Loc	Track	ID	Act	Cap	Intent
First observation $U(G)$	NA	0.84	0.84	0.20	0.52	0.30
Second observation $U(H)$	NA	0.24	0.22	0.52	0.48	0.30
Combined observation $U(K)$	NA	0.62	0.62	0.25	0.40	0.16

NOTE: NA = not applicable.

In either a stochastic or deterministic model, if

$$\sigma_{x1} = \sigma_{y1} = \sigma_1 \text{ and } \sigma_{x2} = \sigma_{y2} = \sigma_2, \text{ then}$$

$$(x_{new}, y_{new}) = \frac{\sigma^2{}_2(x_1, y_1) + \sigma^2{}_1(x_2, y_2)}{\sigma^2{}_1 + \sigma^2{}_2}, \text{ and}$$

$$\sigma^2{}_{new} = \frac{\sigma^2{}_1\sigma^2{}_1}{\sigma^2{}_1 + \sigma^2{}_2}.$$

Once a new location and associated error are calculated from the two observations, it is possible to also calculate the knowledge matrix location column. The entry in the location column is based on the equations described earlier and in the example in Table 2.6. In our example, we took two observations with errors of 10 and 100 units and generated a new knowledge matrix based on the combination of the two observations. The new knowledge matrix shows an 87-percent likelihood that at least 20-unit accuracy is generated in the new observation.

Figure 2.10 illustrates the resulting combination of knowledge matrixes for a stochastic simulation. On the left are the two knowledge matrixes and two uniform random variables associated with the observations being combined. On the right is the resulting knowledge matrix and vector. As we see, the reported quality of the combined observations of the two knowledge types (shown in shaded boxes) is never worse than the best input quality, and one of the knowledge types (activity) is actually better in the combined observation (quality level 5) than in either of the individual observations (quality levels 0 and 2).

The overall quality, as measured by the average quality for each knowledge type, is also higher in the combined observation than that of either of the two input observations individually. For example, even though the reported knowledge level for track ($G_{track} = 1$, $H_{track} = 2$) is no better in the combined observation ($K_{track} = 2$), the average track quality in the combined observation (1.89) is better than that of the two input observations (1.55 and 1.4, respectively).

Figure 2.10
Combining Two Knowledge Matrixes Along with Uniform Random Vectors

Reported likelihood in gray cells.

First observation (G)

Quality	Loc*	Track	ID	Act	Cap	Intent
5	0.118	0.00	0.0	0.00	0.00	0.00
4	0.393	0.00	0.30	0.00	0.00	0.90
3	0.865	0.00	0.80	0.00	0.00	0.90
2	1.0	0.70	0.85	0.00	0.00	0.90
1	1.0	0.90	0.90	0.00	0.90	0.90
0	1.0	0.95	1.0	1.0	0.92	0.90

U(*G*) = (—, 0.84, 0.84, 0.20, 0.52, 0.30)

Second observation (H)

Quality	Loc*	Track	ID	Act	Cap	Intent
5	0.001	0.00	0.70	0.30	0.00	0.00
4	0.005	0.00	0.80	0.40	0.00	0.00
3	0.020	0.00	0.90	0.50	0.70	0.00
2	0.393	0.70	0.95	0.60	0.80	0.30
1	1.00	0.80	0.99	0.70	0.90	0.40
0	1.00	0.90	1.00	0.80	0.92	0.50

U(*H*) = (—, 0.24, 0.22, 0.52, 0.48, 0.30)

Combined observation (K)

Quality	Loc*	Track	ID	Act	Cap	Intent
5	0.119	0.00	0.70	0.30	0.00	0.00
4	0.397	0.00	0.86	0.40	0.00	0.90
3	0.867	0.00	0.98	0.50	0.70	0.90
2	1.00	0.91	0.99	0.60	0.80	0.93
1	1.00	0.98	1.00	0.70	0.99	0.94
0	1.00	1.00	1.00	1.00	0.99	0.95

U(*K*) = (—, 0.62, 0.62, 0.25, 0.40, 0.16)

*Using explicit representation. Mean location quality = 3.38.
RAND *TR416-2.10*

Notice that the quality level generated for the activity knowledge type has increased to 6 even though the individual quality levels were 0 for *G* and 2 for *H*. This large jump in quality points out a drawback of the second approach to combining the uniform random variables. For this reason, the first approach is often preferred. Using that approach, the combined random vector would be $U(K) = (-, 0.24, 0.22, 0.52, 0.48, 0.30)$ and the resulting quality level for activity would be 2. A general rule is to use the second approach when both observations come from the same intelligence domain—single-source fusion—and the first approach otherwise—all-source fusion.

Incorporating Intelligence "Nuggets"

A nugget is a piece of information acquired from a special source or process and usually has unparalleled quality (see Figure 2.11). Nuggets of information combine in the same way as other pieces of information do, and appropriate knowledge matrixes can represent the observation. In some simulations, a special procedure is used for nugget fusion, but this is not required in the knowledge matrix approach.

Figure 2.11
Nuggets May Occur Infrequently, but They Combine Easily

COP

Quality	ID	Act	Cap	Intent
5	0.4	0.3	0.2	0.1
4	0.6	0.5	0.4	0.3
3	0.7	0.6	0.6	0.5
2	0.9	0.7	0.7	0.6
1	0.95	0.8	0.8	0.7
0	1.0	1.0	1.0	0.8

E(*Q*)** = (3.6, 2.9, 2.7, 2.0)

Updated COP has higher mean quality for all knowledge categories

Nugget

Quality	ID	Act	Cap	Intent
5	0.6	0.3	0.2	0.1
4	0.7	0.5	0.4	0.4
3	0.8	0.6	0.6	1.0
2	0.9	0.7	0.7	1.0
1	0.95	0.8	0.8	1.0
0	1.0	1.0	1.0	1.0

E(*Q*)** = (4.0, 2.9, 2.7, 3.5)

Updated COP

Quality	ID	Act	Cap	Intent
5	0.76	0.51	0.36	0.19
4	0.82	0.75	0.64	0.58
3	0.94	0.84	0.84	1.0
2	0.99	0.91	0.91	1.0
1	0.998	0.96	0.96	1.0
0	1.0	1.0	1.0	1.0

E(*Q*)** = (4.5, 3.9, 3.7, 3.8)

*Observations obtained from special sources or processes.
***E*(*Q*) is the mean quality for each column.
RAND *TR416-2.11*

Inference

Table 2.10 shows an example of how to generate additional knowledge of one type from other types of knowledge. This process is known as *inference*: In possession of certain knowledge in a couple of areas, one can infer knowledge in a different area. The criteria for inference in this example have been derived from the Army’s All Source Analysis System (ASAS). There is a separate inference criterion for each type of entity on the battlefield. For example, there might be different tables for airplanes, small and large units, and fixed equipment.

Suppose you have knowledge of track to quality level 3 and location to quality level 2. The ASAS inference tables show that we can infer knowledge for identity to quality level 3, activity to level 2, and capability to level 2. In the cases shown in Table 2.10, the left knowledge matrix (before inference) has a location quality level of 2, a track quality level of 3, and rather poor qualities for the rest of the knowledge types. If we apply the inference criteria, we can infer that such knowledge of location and track will indicate knowledge in identification, activity, and capability. In this example, we set the likelihoods for the inferred knowledge to 90 percent in the matrix on the right. The actual threshold used for the inferred knowledge can be set according to how tightly the information types are coupled. These thresholds are set by SMEs.

Table 2.10
Example of Inference

	Before Inference			After Inference				
Quality	Loc	Track	ID	Loc	Track	ID	Act	Cap
5	0.010	0.00	0.12	0.010	0.00	0.12	0.0	0.0
4	0.030	0.00	0.23	0.030	0.00	0.23	0.0	0.0
3	0.360	0.90	0.39	0.360	0.90	0.90	0.0	0.0
2	0.920	0.92	0.50	0.920	0.92	0.90	0.9	0.9
1	0.970	0.95	0.63	0.970	0.95	0.90	0.9	0.9
0	0.997	0.99	0.80	0.997	0.99	0.90	0.9	0.9

NOTE: In this example, the knowledge matrix before inference has 0 in all other fields.

Aggregates

Aggregates of entities on the ground may be incorporated into the COP in three ways. The first is through direct sensing of a unit. The second is through inference, wherein the observation of a number of individual entities provides knowledge of an aggregate. Both these means of generating aggregate observations are described later in this report. The third method of incorporating aggregate units into the COP is through IPB. In this case, intelligence gathered prior to an operation or campaign has indicated some knowledge of the enemy, and this information has been put into the operational picture to be used in generating the commander's running estimate. The actual units involved might not have been sensed during the operation, but they are still contained in the situation map.

As aggregate entities are collected and integrated into the COP, they can combine with the units generated through the IPB. Because of the nature of the information contained in the IPB, the units may not be handled the same way as other bits of information generated during the campaign. Information contained in the IPB may have much longer hold times, may not age similarly to other entities, and may not be discarded when evaluating criteria for combination of two observations. In this regard, the IPB may be assumed to be very good and a significant factor in the commander's running estimate.

CHAPTER THREE

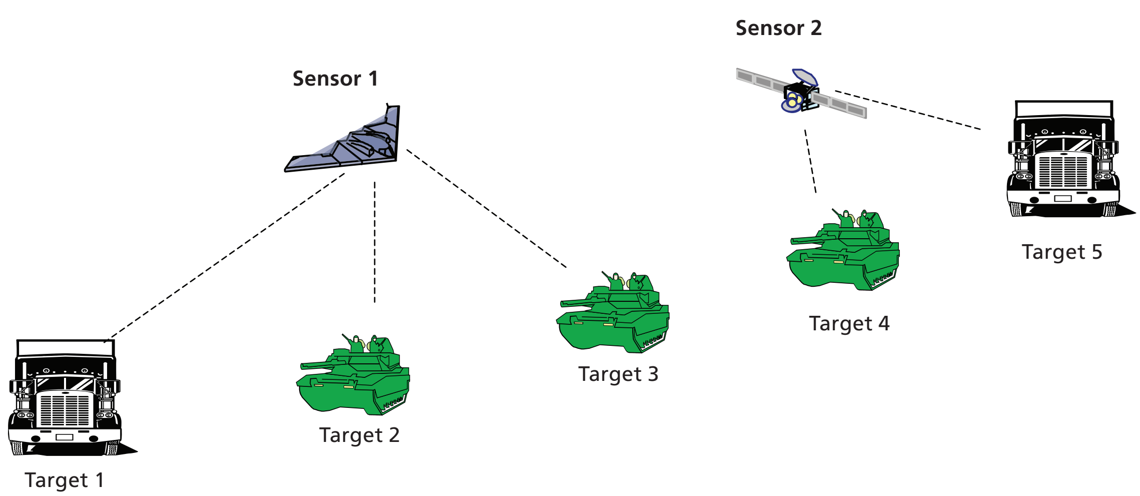
Convoy Example

Figures 3.1 and 3.2 provide a more complex example than the one given in Chapter Two. In this scenario, we have a convoy of five targets arranged equidistantly in a line. The convoy has trucks on the outside positions and three tanks on the inside. Two sensors can observe two and three entities, respectively, at each time step.

The fusion process takes place as follows. We first sort the sightings in the current COP based on their distance to the new observation, with the closest sighting first. Next, we find the first sighting in the sorted list that is close enough to, has consistent identification with, and is good enough to fuse with the current observation. If, during a fusion attempt, only the good-enough criterion is not met, then the observation with the lesser overall quality is removed. It is also possible that sightings in the current COP are removed because of old age. Any remaining sightings in either the new observation or the current COP that have not been previously removed or fused with another sighting are passed on to the next COP.

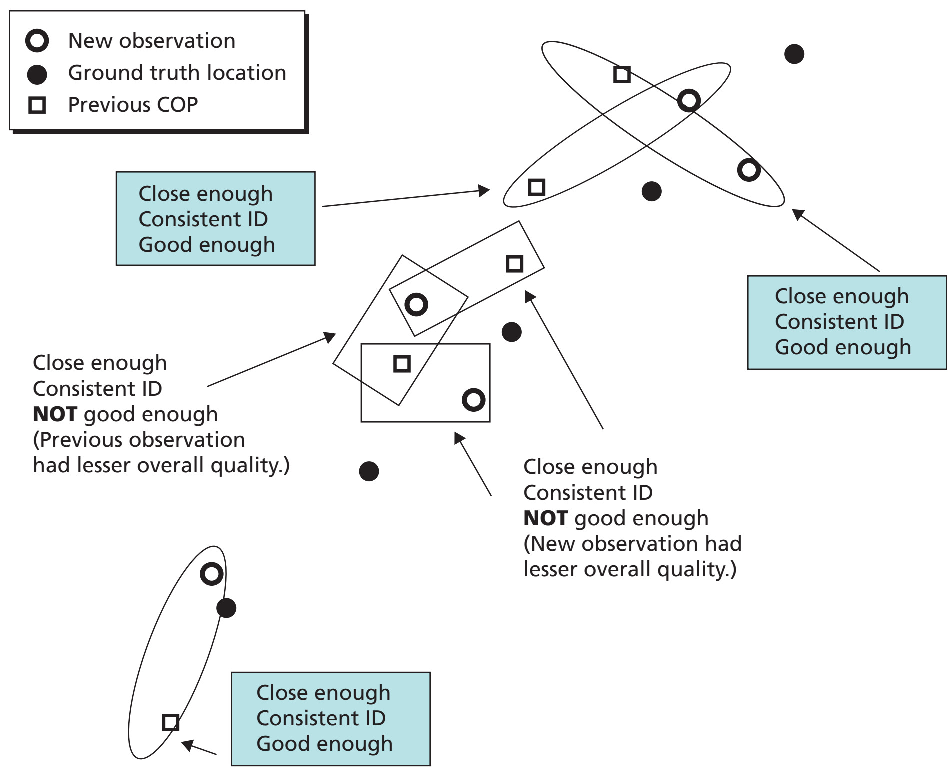
Figure 3.2 shows an example of how the COP is formed at time step 1 using the convoy scenario presented in Figure 3.1. The previous COP observations from time step 0 are indicated by squares, and the new, incoming observations are indicated by circles. The ground truth locations of the five targets are indicated by dots.

Figure 3.1
Complex Example with Three Tanks and Two Trucks in a Convoy



RAND *TR416-3.1*

Figure 3.2
Forming the COP at Time Step 1



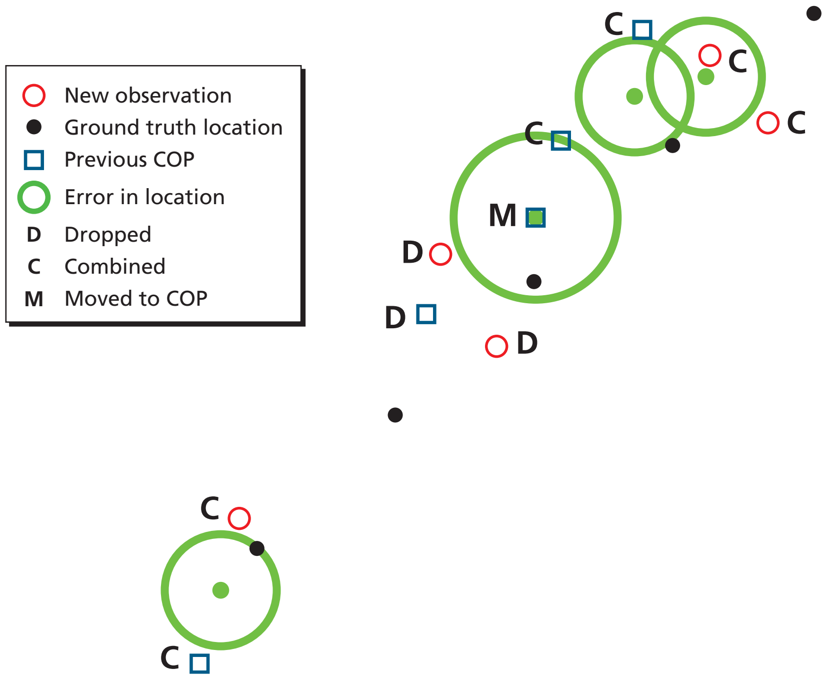
RAND *TR416-3.2*

The rectangles around the pairs of sightings indicate those pairs that failed to fuse because of violation of at least one of the three required criteria. In this time step, three pairs of observations are shown inside the boxes. In each of the cases, the two observations were close enough and had consistent identities, but they were not good enough according to the criteria we established earlier in this report. In these cases, the observation with the lesser overall quality in each pair will be dropped. In the figure, both new and previous observations have been dropped because they are not good enough.

The ellipses around the remaining pairs of sightings indicate those pairs that did meet all three criteria and thus will be combined. Note that, in one circumstance, the closest circle-square pair (of the four contained in the two ellipses in the upper right) are not combined. This may be an artifact of not coming up in the model as meeting the criteria or of the order in which incoming observations are checked against previous observations. Figure 3.3 shows the result of those combinations.

Here, we see the resulting COP at time step 1, shown by the dots within the large circles. The large circles centered at the COP locations denote the circular error probability (CEP) of the observation and correspond to a circle with radius equal to $1.18 \times \sigma$. Next to each sighting from both the previous COP and the new observation is either a D, C, or M to indicate whether this sighting was dropped from the new COP (D), combined (fused) with another sighting (C), or moved (copied) to the new COP (M), respectively. Note that the latter observations (those moved to the COP) are coincident in location with the previous observations and may not be readily seen in the figure.

Figure 3.3
COP at Time Step 1



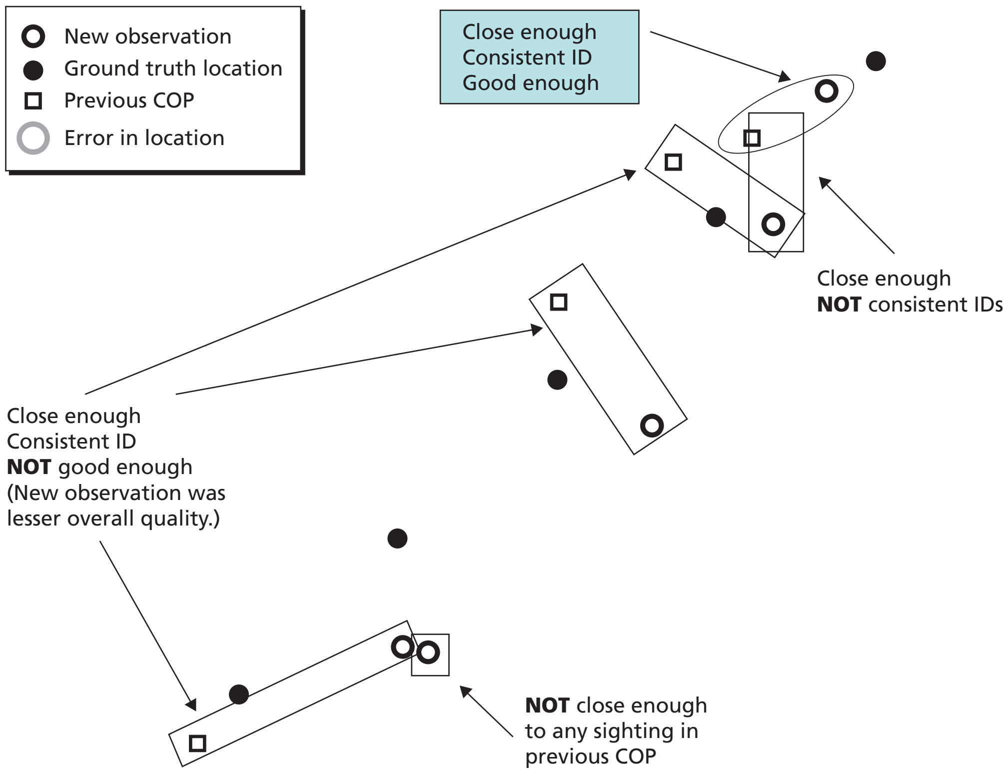
NOTE: Large circles show CEP for fused observation (radius of circle is 1.18 × σ of fused observation).
RAND *TR416-3.3*

Figure 3.4 shows, for time step 2, how the new COP is formed for this same example. In this case, only the pair of sightings in the upper-right corner are close enough, of consistent identification, and good enough to combine. All other pairs fail to pass at least one of these three criteria. In particular, the new observation at the bottom center of the figure is not close enough to any sighting in the previous COP to pass the first criterion. Near the upper-right corner, two sightings are not of consistent identification. The remaining pairs are both close enough and of consistent identification but are not good enough.

In Figure 3.5, the resulting COP at time step 2 is shown by dots, with the location error as denoted by the CEP, shown by the large circles. As in the previous time step, we again label each sighting in the new observation and the previous COP to denote whether it was dropped from the new COP, combined with another observation, or moved directly to the new COP.

Figure 3.6 provides an animated demonstration of a how the COP changes over numerous time steps as new observations arrive. It uses the convoy example with five targets on a diagonal line. The location errors (standard deviations) are approximately one-third of the distance between consecutive targets. At each time step, new observations are fused with the previous COP, dropped, or copied directly to the next COP.

Figure 3.4
Forming the COP at Time Step 2



RAND *TR416-3.4*

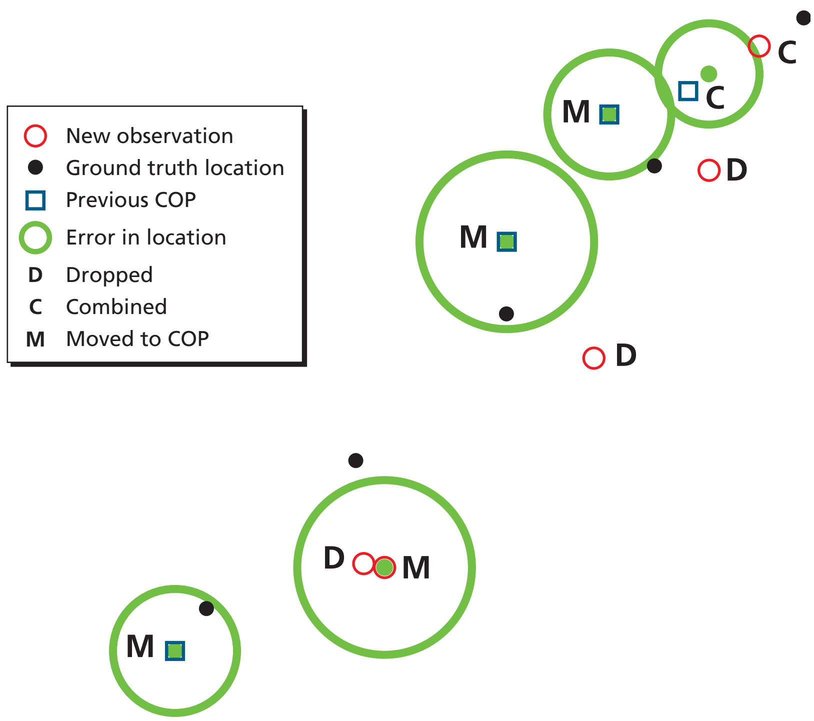
Figure 3.6 shows six consecutive time steps. The first part of each time step shows the locations of the observations on the COP along with the locations of the new observations. The second part shows the results of any combinations that occurred.

The COP in this example contains either four or five unique observations at each time step—close to the total number of observations actually on the ground (ground truth). This may be a result of reasonable errors in location knowledge, which allows the process to discern between two different and two similar entities when new observations are added to the picture.

In addition, over the first five time steps, we see a movement of the COP toward the center line, where the ground-truth entities actually are. At time step 6, however, two observations show up toward the bottom left of the convoy, and a blank area is evident in the COP near the middle of the convoy. The former may be the result of a combination that did not happen as a result of a stochastic draw that did not favor combination (even though it seems that they should have been combined). The latter may be a result of dropping an observation on the COP that had aged past its threshold.

Using the same scenario used in Figure 3.6, Figure 3.7 uses animation to show how the COP changes over time. Each frame shows the ground-truth locations of the targets, as well as the current COP and errors in their locations as given by the corresponding CEP circle. Over time, we see that the COP moves closer to the ground-truth locations as additional intelligence is built up. In addition, we see that the errors are reduced as more observations are added to the collective knowledge of each entity.

Figure 3.5
COP at Time Step 2



RAND *TR416-3.5*

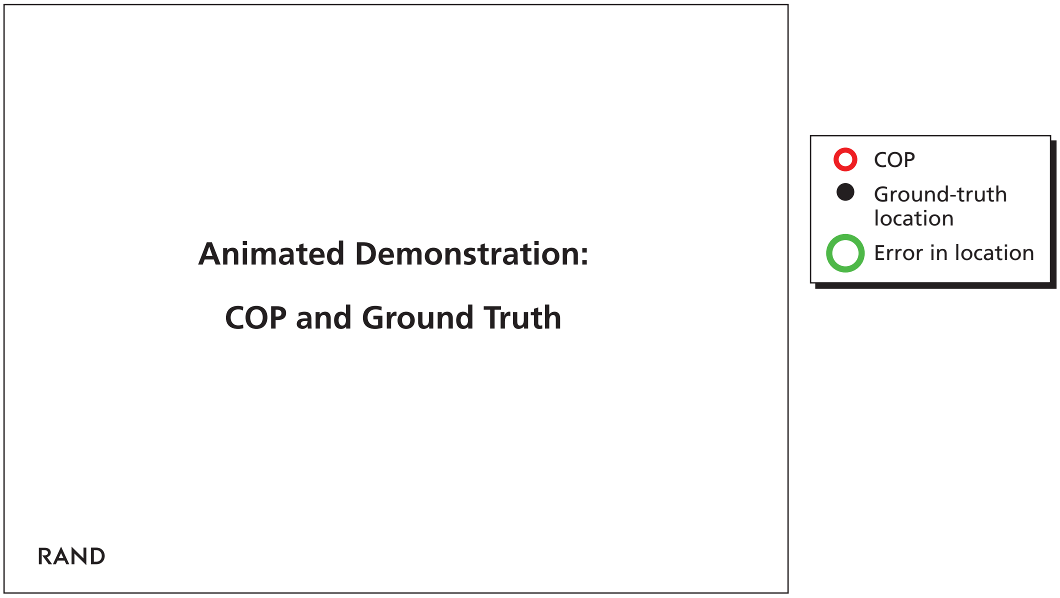
Figure 3.8 is similar to Figure 3.7 in that it uses animation to show the COP changing over time for the same scenario, but Figure 3.8 includes location errors (standard deviations) that are three times that of the previous example. As a result, we have numerous extraneous observations on the COP that are simply added to it from one time step to the next, rather than being combined with existing sightings. Not surprisingly, the resulting COP is further from ground truth than that in the previous example, due to the increased location errors. Indeed, the location errors have decreased the quality of the COP.

Figure 3.6
Populating the COP

This demonstration shows changes in the COP over six consecutive time steps with the addition of new observations using the convoy example with five targets on a diagonal line. (To view, please click on the demonstration.)

RAND *TR416-3.6*

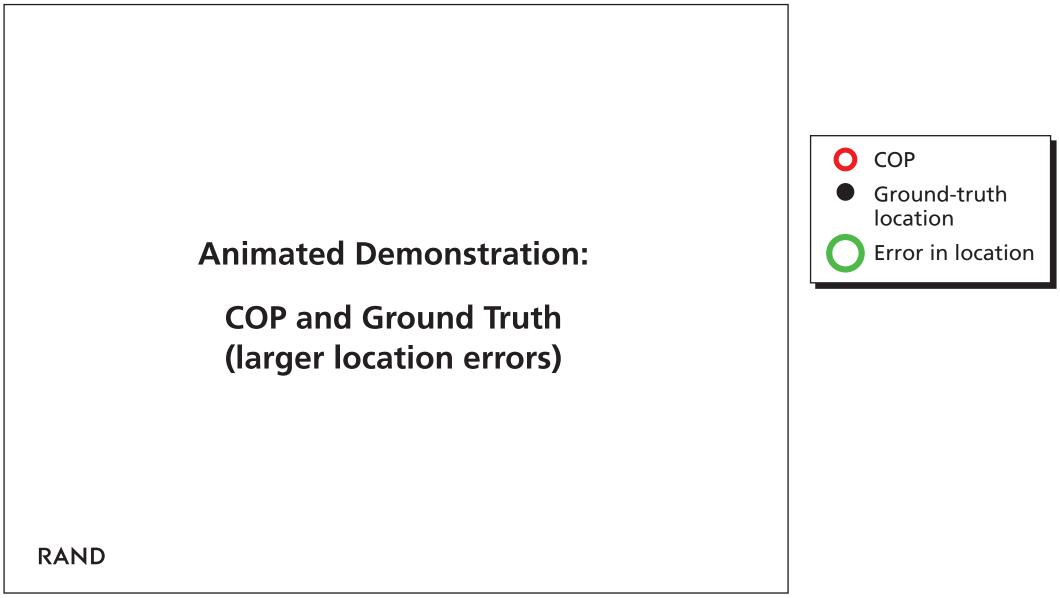
Figure 3.7
Changes in the COP Over Time



Using the convoy example with five targets on a diagonal line, this demonstration shows how the COP and location errors change over time with respect to the ground-truth locations. (To view, please click on the demonstration.)

RAND *TR416-3.7*

Figure 3.8
Increased Location Errors Result in Extraneous Observations on the COP



This demonstration uses the convoy example with five targets on a diagonal line to show changes in the COP with increased location errors, degrading the quality of the observation. (To view, please click on the demonstration.)

RAND *TR416-3.8*

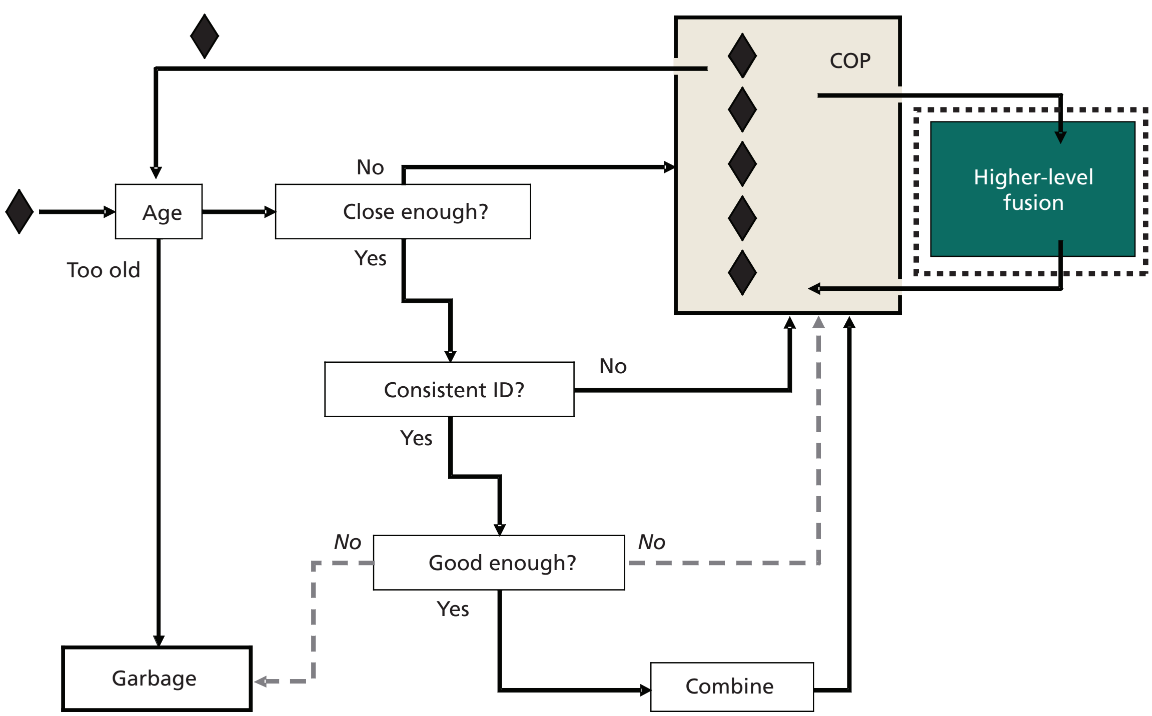
CHAPTER FOUR

Higher-Level Fusion

This chapter describes some representations of higher-level fusion—namely, fusion levels 2 and 3. Higher-level fusion can be thought of as part of the overall process of populating the COP (see Figure 4.1). Unlike the process of lower-level fusion, in which the COP is populated solely by observations coming into the system, in some of the higher-level fusion processes, the COP is populated by the operations performed directly on itself. Here, we discuss the generation of aggregates and higher-level units on the COP.

The output from level 1 fusion is a set of discrete, observed battlefield entities with information about the type, location, movement, identity, status, and capability of each entity, along with an evaluation of the quality of this information. Based on the level 1 products, IPB OOB templates, and knowledge about the environment, level 2 fusion aggregates discrete

Figure 4.1
Steps Involved in Populating the COP: Higher-Level Fusion



NOTE: Diamonds signify observations.
RAND *TR416-4.1*

entities into larger objects that are interacting. It interprets entity events and actions and hypothesizes what events may occur next. The outputs from level 2 are aggregated, as are inferred entities, observed and derived events and actions, and a collection of hypotheses regarding events that might happen in the future. In addition, quality assessments of these products will be available. Level 3 fusion projects a current situation into the future to predict intent and courses of action.

There are three methods of detecting an aggregate. The first addresses the direct observation of an aggregate by a sensor. The next two methods are indirect methods of detecting an aggregate, either through the observation of one entity and inferring its parent or through observations of many entities, implying a larger collection. We describe each of these three methods.

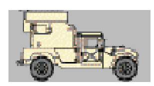
One means of detecting an aggregate is through direct observation of the unit with a sensor (see Figure 4.2). Some sensors are much better at identifying aggregates than individual entities and, thus, can directly observe an aggregate without having "seen" the entities that make up the aggregate. The SIGINT sensor is such an example: Nodal analysis can produce a direct observation of an aggregate unit. The ability to make this observation is dependent on the processing time for the analyst to determine whether a unit has been seen and the probability that a certain type of unit can be discerned with the sensor.

An aggregate is a collection of units. For instance, a battalion is an aggregate of multiple companies, and a company is an aggregate of multiple platoons. Sometimes, if enough information is obtained on a unit, its parent may be determined. For example, with enough knowledge of the identity of a tank, the platoon of which it is a part may become evident. Likewise,

Figure 4.2
Sensors Can Generate Both Entity and Aggregate Detections

If a number less than 0.7 is sampled, an observation of a parent unit is generated with its own appropriate knowledge matrix.

Quality	ID
5	0.70
4	0.80
3	0.90
2	0.95
1	0.99
0	1.0



Quality	Loc*	Track	ID	Act	Cap	Intent
5	0.118	0.00	0.0	0.00	0.00	0.00
4	0.393	0.00	0.30	0.00	0.00	0.90
3	0.865	0.00	0.80	0.00	0.00	0.90
2	1.00	0.70	0.85	0.00	0.00	0.90
1	1.00	0.90	0.90	0.00	0.90	0.90
0	1.00	0.95	1.0	1.0	0.92	0.90

RAND *TR416-4.2*

knowing enough about the identity of a platoon may allow inferred knowledge of the battalion to which it is subordinate.

The knowledge of identification is captured in the identification column of the knowledge matrix. A quality level of 5 in the identification column implies enough knowledge to determine the identity of the parent unit. In the case in Figure 4.2, a random draw of less than 0.7 would generate an observation of the parent in addition to the unit that has been seen. The parent would similarly have a knowledge matrix and vector associated with it from the sibling detection. This is one means of detecting aggregates. A final method of aggregate detection is the direct observation of an aggregate or some portion of it. We describe this method next.

Aggregate detections can sometimes be generated by a group of entity detections. To infer an aggregate from a group of entities, it is necessary to have templates for the various types of aggregates that may be present.[1] These templates, which define both the number and types of entities that make up a specific aggregate, are defined by SMEs.

There are two methods by which an aggregate may be inferred from a group of entities. If a minimum percentage (defined by SMEs) of subordinate units is detected, then the superior unit (aggregate) is considered detected. This inference of the aggregate is an integral part of the planning process and helps to generate the commander's running estimate of the enemy forces. An alternative but similar approach is to infer an aggregate by the type of subordinate entity. In other words, if a certain minimum percentage of each type of entity is detected within a given unit, then the superior unit is considered detected.

---

1 These templates are generated as part of the IPB task.

CHAPTER FIVE

Conclusions

The process of fusion, combining pieces of information to produce higher-quality information, knowledge, and understanding, is often poorly represented in constructive models and simulations that are used to analyze intelligence issues. However, there have been efforts to rectify this situation by incorporating aspects of information fusion into combat simulations. This report describes one approach to capturing the fusion process in a constructive simulation, and it provides detailed examples to aid in further development and instantiation.

The method is a sequential process of determining the probability that two pieces of information generated on the battlefield are either from the same battlefield entity or are two separate observations. The process entails putting the observations through a sequence of operations to determine whether they (1) are close enough geographically with respect to their separate errors in location to be of the same entity, (2) have consistent identities that would not prevent them from being considered the same entity, and (3) contain information content of high enough quality to warrant the combination.

Once two observations have passed these three tests, a combination process determines the fused product and passes it on to the COP. In cases in which additional information about an entity is generated—for example, knowledge of the location of a superior unit gleaned from knowing where the subordinate is—the process is able to capture the information in the COP. Higher-level fusion processes, such as the generation of aggregates, are also captured in the fusion process.

In this report, we have described a method for incorporating a representation of fusion levels 0 through 2 into a constructive simulation. The method incorporates both explicit and implicit representations of the fusion of intelligence. The method as described is model independent, and it may be integrated into other Army models as is seen fit.

The method described in this report is a new, nonparametric approach to the representation of intelligence fusion for simulations. Most other approaches use ground truth, are explicit (i.e., they try to replicate the actual procedures), or are not documented. To our knowledge, the approach herein may be the first workable approach that permits both explicit and implicit representations of intelligence fusion and provides the execution speed and flexibility to be easily incorporated into Army simulations.

During the course of this research, we were challenged by the representation of higher-level fusion. For instance, level 3 (predicting future situations) remains a challenge to operational practitioners and will continue to advance in simulations as research advances.

The method described in this report is data intensive: Sensor and collector attributes need to be represented in terms of the knowledge types in the method and probabilities calculated from experiments or available data. Discussions with Army experts have led us to

believe that the collection and use of such data is feasible, though the capability will need to be developed.

References

Bedworth, M., and J. O'Brien, "The Omnibus Model: A New Model of Data Fusion?" *IEEE Aerospace and Electronic Systems Magazine*, Vol. 15, No. 4, April 2000, pp. 30–36.

Blasch, Erik P., and Susan Plano, "Level 5: User Refinement to Aid the Fusion Process," in Belur V. Dasarathy, ed., *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2003* (proceedings of the SPIE), Vol. 5099, April 2003, pp. 288–297.

Dasarathy, B. V., "Sensor Fusion Potential Exploitation—Innovative Architectures and Illustrative Applications," *Proceedings of the IEEE*, Vol. 85, No. 1, January 1997, pp. 24–38.

Endsley, Mica R., "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors*, Vol. 37, No. 1, March 1995, pp. 32–64.

———, "Theoretical Underpinnings of Situation Awareness: A Critical Review," in Mica R. Endsley and Daniel J. Garland, eds., *Situation Awareness Analysis and Measurement*, Mahwah, N.J.: Lawrence Erlbaum Associates, 2000.

Headquarters, U.S. Department of the Army, *Operations*, Washington, D.C., Field Manual 3-0, June 14, 2001. As of July 3, 2007:
http://www.dtic.mil/doctrine/jel/service_pubs/fm3_0a.pdf

Keithley, Hans, *Multi-INT Fusion Performance Study*, Joint C4ISR Decision Support Center, DCS-00-02, 2000.

Kenney, J. F., and E. S. Keeping, *Mathematics of Statistics,* Part Two, 2nd ed., Princeton, N.J.: Van Nostrand, 1951.

Salerno, J., "Information Fusion: A High-Level Architecture Overview," *Proceedings of the Fifth International Conference on Information Fusion*, Annapolis, Md., June 8–11, 2002, Institute of Electrical and Electronics Engineers (IEEE), June 2003.

Shafer, Glenn, *A Mathematical Theory of Evidence*, Princeton, N.J.: Princeton University Press, 1976.

Steinberg, Alan, and Nick Rosen, *A Performance Model for Track Generation, Loss and Recovery of Surface Targets*, Arlington, Va.: TASC, Inc., December 5, 1997.

U.S. Army Intelligence Center, Directorate of Combat Development, "Fusion: An Operational Assessment," Ft. Huachuca, Ariz., July 6, 2004.

Washburn, Alan, *Location Fusion in Land Combat Models*, Monterey, Calif.: Naval Postgraduate School, NPS-OR-05-003-PR, December 2004. As of July 3, 2007:
http://stinet.dtic.mil/cgi-bin/GetTRDoc?AD=ADA429359&Location=U2&doc=GetTRDoc.pdf

Zelen, Marvin, and Norman C. Severo, "Probability Functions," in Milton Abramowitz and Irene A. Stegun, eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th printing, New York: Dover, 1972, pp. 925–996.